# Spatio-temporal organization of one's personal image collection with model-based ICL-clustering

Antoine Pigeau, Marc Gelgon

# SPATIAL-TEMPORAL ORGANIZATION OF ONE'S PERSONAL IMAGE COLLECTION WITH MODEL-BASED ICL CLUSTERING

*A. Pigeau, M. Gelgon*

Institut de Recherche en Informatique de Nantes, INRIA ATLAS group
2, rue de la Houssinière - BP 92208, 44322 Nantes cedex 03 - France
E-mail:surname@irin.univ-nantes.fr

**ABSTRACT**

This paper addresses the issue of automated organization of a personal image collection, in particular to respond to the emerging needs from a mobile camera phones. The issues related to browsing through large image collections acquired from such devices are first discussed. In contrast with metadata-less collections, which necessarily rely on image content, we propose a collection organization technique based on picture geolocation and timestamps. These are indeed available and generally reliable in the proposed context. The objective is formulated as an unsupervised classification problem, in both space and time. The statistical integrated completed likelihood criterion is chosen, providing effective solutions both to model complexity determination and the cluster separability objective, in a setting which avoids arbitrary algorithm parametrization. Reliability of space and time partitions obtained are then assessed, to select an effective segmentation, which may then provide a calendar-type structured view for navigating in the picture collection.

## 1. PROBLEM AND CONTEXT

Content-based image retrieval problems have been dealt with for the past few years, and the field is now equipped with both recent contributions [14] and surveys [23]. It was recently stressed in [1] that extending the scope of this field to address personal image collections was an important stake, and that there was little existing. Very recently, proposals have emerged from most leading industrials in the field of multimedia [20, 18, 19, 10, 16], as well as from academics [13, 15].

We advocate that, within personal image collections, an interesting niche is emerging. Compared with workstations, or even digital cameras, *camera phones* are always carried by their user, and are hence advantageous both as an image acquisition device and as an image retrieval terminal [24]. Indeed, despite the technical issues image retrieval sets on such devices, the permanent availability and the ability to easily share retrieved pictures (eg. through Multimedia Mobile Messaging) makes it appealing.

As images are gathered, they progressively builds up a valuable memory of one's life, which can be later searched, whether for practical, emotional, or pass-time purposes. The system should offer users not only the possibility of quickly, reliably and comfortably retrieve a well-defined piece of information in their potentially large collection, but also functions for browsing, simply to get an overall idea of the content of the collection. Providing such overviews is actually also beneficial, let alone necessary, for retrieving a well-defined piece of information.

A survey of technical issues pertaining to camera phones is provided in [19], while insight in their usage may be found in [24]. One of the conclusions drawn in [19] is that the handling (storage and retrieval) of image collections in this context is an important industrial need and overall an open problem. Furthermore, we believe this issue is both well-defined (probably better than all-purpose content-based image retrieval), and is of practical significance in the years to come.

Personal image collections may be distinguished from the ordinary "digital library" viewpoint by :

- the content itself (nature of the scenes, structure of the image collection, attached meta-data) ;
- the partial memory that the user has of the collection : the user does not discover a wholy new collection, but progressively recalls his past as navigation progresses, thus determining incrementally directions in which (s)he should browse. Browsing presents advantages over querying to this respect ;
- the desired search/browse criteria, as detailed below.

With regard to the latter point, user studies [21, 22] suggest that the most convenient browsing criteria are time, geographical location, and "image annotations/semantic image content/topic". The identity of people present is also among appealing criteria. On the other side, classical features such as color, shape, layout and texture cues are rated of little relevance.

Research in content-based image retrieval has often been justified by the lack of meta-data. With camera phones, both time and location measursements may be assumed, corresponding to very criteria that users wish to use.

The focus of the present paper is the generation of a structured representation of the image collection, allowing the user to easily browse through time and space. In this paper, we solely consider time and geolocation meta-data attached to each image, the image content itself is ignored. Further, we wish to make the scheme as unsupervised as possible, i.e. the temporal and spatial bounds and extent of the image groups, and the number of these groups, should be determining from the data.

Finally, human-computer interactions considerations are important, as they may define or affect content analysis goals. Research contributions that have been put forward mainly target the workstation/ebook context, and propose a variety of strategies for laying out image sets [15]. In the case of small mobile devices with stringent input and display constraints, hardly more than one to three image may be displayed simultaneously. We argued in [11] that a crux is the ability to generate summaries according to the criteria identified above as relevant, so as to suit visualization and browsing needs. The present work proposes a contribution in this direction.

The remainder of this paper is organized as follows. Section 2 surveys existing proposals that exploit time or space for retrieval in image collections. In section 3, we present the proposed technique for spatio-temporal organization of one's image collection, first as an overview and then in more detail. Section 4 provides experimental results, while section 5 depicts a typical user interface exploiting the results. Finally, section 6 is devoted to concluding remarks.

## 2. RELATED WORK

### 2.1. Time-based structuring

Structuring an image collection according to the time stamp of each picture is intuitively appealing, practically quite cheap and reliable. As noted in [13], the generative process of pictures (i.e. behaviour of users) is likely to exhibit time clusters and, furthermore, often in a hierarchical fashion. Overall, two types of techniques may be distinguished. First, change detection techniques, such as in [20], possess the advantage of not setting a particular parametric model on the intra-cluster time distribution. A combination of these alternatives is proposed in [13], in which (preset size)-gap detection leads to initial groups for clustering. However, how the classical limitations of clustering techniques are addressed is not detailed (number of clusters, arbitrary intra/inter-class separation thresholding). In order to cope with the variety of time scales present in the image collection, solutions such as log-scaling of inter-frame time gaps have been examined in [20, 16]. Finally, besides direct use of time for image grouping, it was recently proposed [10] to combine time linearly with camera settings features and image content information, within an 'image similarity' measure.

### 2.2. Geolocation-based structuring

As recently reviewed in [25], geolocation technologies are progressively being integrated in mobile phones and networks. In practice, geolocation is mainly pushed on the market by navigation and context-aware services [17], rather than our image retrieval purpose, but we nevertheless benefit from it and it is of utmost importance.

The importance of the location for image collection organization is stressed in [13] but, to our knowledge, there are currently few systems that seem to have considered the matter closely. Regardless of image consideration but still with a view to providing structured calendar-type views, we proposed in [12] a technique towards unsupervised learning of meaningful locations. Geolocation is measured continuously in time, and partitions of time and space are extracted at multiple scales, based on a piece-wise parametric trajectory model. By this means, one attempts to recover significant temporal episodes and areas. A work close in spirit is [2], although the modelling formalism differs.

## 3. SPATIO-TEMPORAL ORGANIZATION WITH MODEL-BASED ICL CLUSTERING

### 3.1. Meta-data used and overview of our proposal

The objective is to create a structured representation of the image collection, allowing the user to easily browse through time and space. In this paper, we solely consider time and geolocation meta-data attached to each image, the image content itself is ignored. Further, we wish to make the scheme as unsupervised as possible, regarding the temporal and spatial bounds and extent of the image groups, and the number of these groups.

We formulate this goal as a model-based unsupervised classification problem. Here are the main features of the proposed scheme :

- distinct classifications are built for time and space, as illustrated by fig. 1 ;
- we resort to the statistical integrated completed likelihood criterion, initially proposed in [3]. Indeed, with its evidence-like aspect, it provides an effective solution to model complexity determination, i.e. determining a suitable number of clusters. Besides, it drives the search towards favouring cluster separability, to provide partitions that are more exploitable for our purposes. A practical advantage, important for our application, is some degree of robustness to mismatch between the (Gaussian) form of probabilistic model components and the actual clusters ;

- optimization of this criterion is conducted using an Expectation-Maximization technique, with a dedicated search procedure ;
- the spatial and temporal classifications found are assessed, with respect to how well they succeed in separating the data into clear groups. The partition retained may then provide a structured, calendar-type (or map-type) organization and view of the picture collection.

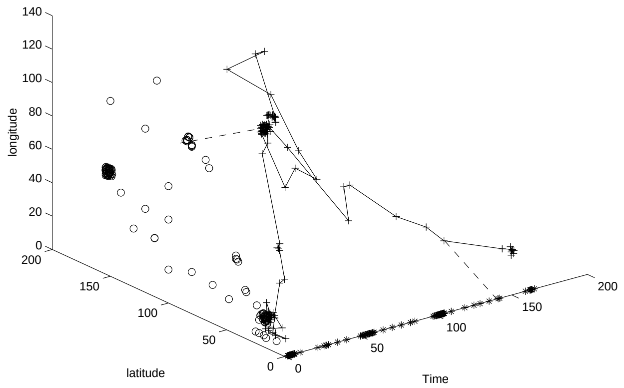



**Fig. 1**. Projection of an example data on time (*) and geolocation (o) axes. The line represents the spatio-temporal displacement of the user, the '+' indicate places where&when photos are taken. Let us underline that clustering is not conducted directly in the three-dimension (x,y,t) space.

### 3.2. Optimality criterion

Model-based clustering is a favourite framework for identifying meaningful groups in data [9]. Overall, it requires setting a functional form on the probability distribution of the data arising from each cluster, defining a statistical optimality criterion and searching for a suitable solution accordingly.

By taking a Bayesian hypotheses testing viewpoint, it can be shown that an effective manner of evaluating the ability of a clustering hypothesis $H_i$ to explain the data $D$, taking into account the need for comparing hypotheses with various numbers of clusters, is provided by the so-called *evidence*, or marginalized likelihood :

$$P(D|H_i) = \int P(D|w_i, H_i)P(w_i|H_i)dw_i \qquad (1)$$

where $w_i$ indicates the model parameter vector associated to hypothesis $H_i$.

The goal is thus to find the mixture model leading to the greatest evidence for clustering the data. In order to evaluate this marginalized likelihood, a variety of computations and approximations exist, as reviewed in [6]. We opt here for the approximation known as the Bayesian Information Criterion (BIC) [9]. expressed as follows :

$$BIC = -ML + \frac{1}{2} \cdot N(K) \cdot log(n) \qquad (2)$$

where $ML$ is the maximized mixture loglikelihood, $N(K)$ is the number of independent parameters in the model with $K$ components and $n$ is the number of data elements. An intuitive observation on this approximation is that it acts as a likelihood criterion penalized by model complexity, illustrating the use of marginalized likelihood as the Bayesian implementation of Occam's razor.

Overall, the BIC criterion aims at identifying both model parameters and the number of clusters. Yet, the parametric form enforced for clusters (henceforth, in practise, Gaussianity) leads to poor and/or over-segmented clusters, should their data strongly infringe on this assumption. In our case, groups of pictures taken in meaningful areas or time extents do not generally exhibit very Gaussian distributions, as their physical generative process is not so.

To improve this point, we resort to the Integrated Classification Likelihood (ICL), proposed in [3]. This criterion may be summarized as follows :

$$ICL = BIC - \Phi(K) \qquad (3)$$

where $\Phi(K)$ is the estimated mean entropy of the mixture, defined by :

$$\Phi(K) = -\sum_{k=1}^{K}\sum_{i=1}^{n} t_{ik} log t_{ik} \geq 0 \qquad (4)$$

where $K$ is the number of components in the mixture, $n$ the size of the data set, and $t_{ik}$ is the posterior probability for an observation $i$ of originating from cluster $k$. These $t_{ik}$ values are supplied at convergence of the optimization phase, which is described in the next section.

Overall, this criterion penalizes the BIC criterion with the entropy of the data-to-model assignment in the mixture, i.e. a mixture which separates well the data has lower entropy and is favoured.

In the next section, we present the technique for attempting optimization of this criterion.

### 3.3. Search for the optimal classification

The Expectation-Maximization (EM) algorithm [7, 5] is used for (locally) optimizing the criterion described above. This algorithm is widely used in association with mixture modelling. It iterates between two steps, the E-step, in which the unknown data-to-model assignements are replaced by their expectation, given the current model parameter estimates, and the M-step in which the model parameters are estimated, given the current data-to-model assignements.

We describe hereunder several problems the EM algorithm has to cope with, and some solutions proposed in our case.

– As the EM technique assumes the number of components is known, one has to conduct the search for each hypothesis subspace associated to a given number of components. Among all the solutions found, one retains the solution which maximizes the ICL criterion. An improvement on this exhaustive search strategy is proposed in [8]. However, we aim at displaying the partitions on the screen of a small-size device, which provides us with some upper bound on the number of clusters ; hence, we carry out optimization exhaustively between 2 and 15 components (the 1 cluster option is considered but does not require optimization, only evaluation of the ICL criterion). This is also motivated by the interest in providing partitions at multiple granularities, i.e. finding several plausible partitions with various complexities, to enable multi-scale browsing. This is left for future work.
– It breaks down when there are too few observations associated to a component. For instance, in the case of a component with a single observation, the variance of a component is undefined. Still, it is common to find a cluster consisting of an isolated data element. To alleviate this, we propose to enforce a minimum variance in time (or covariance in space) for each cluster, at each M step. For the geolocation-based clustering, for instance, the covariance matrices are computed as follows :

$$\Sigma_{M_{step}} = \begin{cases} \Sigma_{ML} & \text{if } |\Sigma| > \epsilon \\ \begin{pmatrix} \sqrt{\epsilon} & 0 \\ 0 & \sqrt{\epsilon} \end{pmatrix} & \text{otherwise} \end{cases} \tag{5}$$

where $\Sigma_{ML}$ indicates the max. likelihood variance estimate and $\epsilon$ is a small minimum variance.
This technique does not set a troublesome assumption, as time between two pictures or the accuracy of the geolocation sensor sets anyway a lower bound on the size of meaningful clusters ;
– the third issue is due to the local character of the optima found with the EM algorithm. Instead of the commonly used k-means-based initialization, we resort to a procedure that seems more effective [4]. This so-called "em-EM" technique, consists in running several times the following set of sucessive steps : randomly initialized k-means, followed by a short number of EM iterations. This provides a set of candidate points for further search, through many more EM iterations. Overall, this technique appears to provide a reasonnable compromise between computation time and the quality of minima found. In the time domain, the set of random k-means initializations is constrained to form connected temporal components, thus generally starting closer from desirable solutions.

### 3.4. Comparison of spatial and temporal partitions

Let us recall that the classifications are carried out independently in space and time. Due to the lack of natural clusters in the data, to insufficiency of the optimality criterion or poor local minima, one or both of the partitions obtained would occasionnally poorly capture the data structure. We propose to compare the values of entropy $\Phi(K)$ found for each of the two optimal classifications, and select the one with the smallest value, i.e better-defined clusters. This generally leads to selecting the most relevant and reliable partition for viewing and browsing the picture collection.

## 4. EXPERIMENTAL RESULTS

The EM algorithm is initialized with 500 em-EM short runs (a maximum of 20 iterations for each run), followed by a maximum of 500 EM iterations from the best solution obtained on short runs.

The first experiment reported corresponds to a one-day touristic walk, during which photos at taken at interesting sights. The user starts at location A (cf. fig.2), goes to B, then C and returns to A. The classifications obtained in time and space (along with ICL-based comparison of models of various complexities) are provided respectively on fig. 2A and fig. 2B. For comparison sake, a hand-made "ground truth" built by an an external person is supplied below, in fig. 3.

$5$ and $11$ classes are obtained in, respectively, the time and the geolocation partitions. On the temporal axis, there seems to be actually little cluster structure in the data. On the geolocation side, there exists meaningful locations, which are quite correctly identified, but minor errors still remain. For example, location C is divided in two classes. Exhaustive combinatorial search was conducted to determine whether such residual problems were due to modelling and criteria issues, or if it they correspond to suboptimal local minima. It comes out that a large majority are due to local minima. Although this is effectively improved by increasing the number of em-EM short-runs, we are starting evaluating better quality/cost multi-scale optimization techniques. The intent is two-fold : providing a representation of the data at multiple scales, for viewing/browsing, but also using this to drive the optimization procedure.

The separability of the data in time and space is assessed as $\Phi_{time} = 4.55$ and $\Phi_{location} = 5.29$. These values are of the same order of magnitude ; as the time-based entropy is smaller, the time-based view is selected. This is also related to the fact that, in this very example, time-based picture groups in fact correspond quite well to location-based groups, hence deciding on the time or space criterion on little grounds has little effect. More generally, as one would expect, it can be observed the entropy $\Phi$ is tightly correlated to the mismatch between manual ground truths and classifications obtained automatically.

We now focus on two other scenarios. In practice, they are obtained by considering subsets of the global scenario, first focusing on zone B, then on zone A. However, all three classifications are carried out independently. We first consider area B, corresponding to a period when the user is visiting a famous square. The classifications obtained on the temporal and spatial domains are presented in fig. 4. The photos, which are quite evenly distributed in time, are found to be best "clustered" into a single class. This is not a practical option for summary-based calendar presentation of the collection. On the other side, the geolocation partition, which computed classification matrix entropy mean entropy is evaluated as $\Phi_{location} = 0.01$, hence well-clustered, is chosen. We did not accurately examine the case where one should decide between a single class, or an unreliable partition, characterized by high entropy, as this would require more user-based studies.

Figure 5 presents the same type of experiment, on zone A. The classifications obtained are composed of $5$ location-based components or $2$ time-clusters. In this case, the time partition is highly structured ($\Phi_{time} = 0.01$) and hence rated more relevant the location-based one ( $\Phi_{location} = 6.31$).

Although experiments were so far conducted on data sets corresponding to the scale of days rather then globally to weeks and months, they provide a first validation of the proposed scheme.

## 5. APPLICATION TO IMAGE COLLECTION BROWSING

In this section, we illustrate the practical interest of the proposed scheme, from an image collection browsing perspective, as applied to a mobile phone/PDA.

We consider here time-oriented views, such as would be found on familiar calendar managers available on PDAs. The proposed divisions of time for browsing may be either the temporal or spatial partition, depending on the above-mentioned criterion. In the case space is selected, temporally disconnected components are separated. Restriction to one of the partitions ensures a small number of episodes and consistency of the division criterion along the view. The choice of suitable representative images for image groups is out of the scope of this paper, and some solutions are proposed in works referred to in section 2.

Alternatively, one may exploit both partitions simultaneously. Fig 6 illustrates a possible calendar-type view. Different background colors or separation marks between episodes indicate from what perspective the cluster is homogeneous. Input keys can be associated to the following functions " jump to next temporal episode","jump to the next occurrence of photos in this spatial zone".

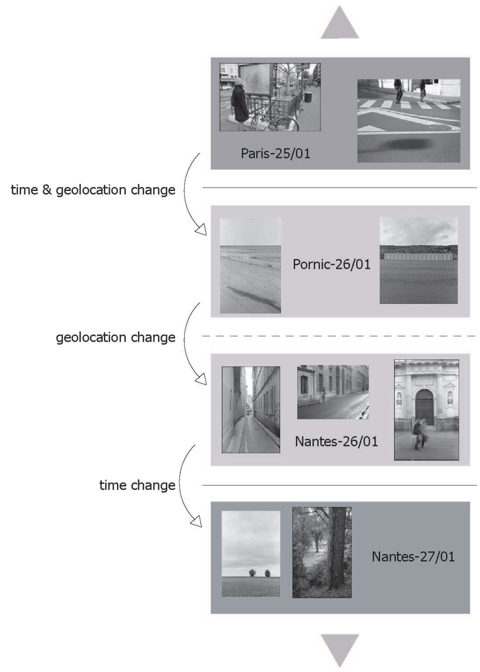



**Fig. 6**. Example of electronic calendar : dashed lines represent a change of geolocation, and continuous lines a time change or both. Assignement of names to clusters is carried out manually and does not relate to the proposed technique.

## 6. CONCLUSION

In this paper, the problem of organizing personal digital image collections has been addressed. The interest towards temporal and spatial meta-data was stressed, especially for camera phone applications. To focus on purely data-driven structuring, we formulated the problem as an unsupervised classification issue. The ICL criterion was proposed, as it fulfills several requirements of the application (unknown number of clusters, poor Gaussianity). It is optimized with an efficient version of the EM technique. This ICL criterion was further used for selecting the more relevant between the temporal or spatial partitions. Among the qualities of the chosen formalism are the lack of delicate parameter tuning, and perspectives for multi-scale structuring. Overall, we believe it provides an effective and realistic direction towards organizing an image collection, providing structure may be mapped onto a variety of navigation schemes.

In current work, we are examining lowering the cost of the optimization strategy. More precisely, this should be carried out jointly with the structuring of the collection at multiple scales. Among other perspectives, turning to an incremental system is essential. Although the EM algorithm adapts quite well to this case, more global reorganization of the collection sometime has to be carried out.

## 7. REFERENCES


[1] P. Anandan. Personal digital media : It's about sharing experiences. In *MMCBIR 2001 - Multimedia Content-based Indexing and Retrieval*, INRIA Rocquencourt, France, Sep. 24–25 2001.

[2] D. Ashbrook and T. Starner. Learning significant locations and predicting user movement with GPS. In *IEEE Int. Symp. on Wearable Computing (ISWC'2002), Seattle, USA*, Oct. 2002.

[3] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated classification likelihood. In *IEEE Transaction on PAMI*, volume Vol.22, pages 719–725, Jul. 2000.

[4] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics and Data Analysis*, pages 561–575, 2003.

[5] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Nov. 1995.

[6] D.M. Chickering and D. Heckerman. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. Technical Report MSR-TR-96-08, Microsoft Research, March 1996.

[7] A. P. Dempster, N. M. Laid, and D. B. Rubin. Maximum likelihood for incomplete data via the EM algorithm. *J. R. Stat. Soc.*, pages 1–38, 1977.

[8] M. Figueiredo and A. K. Jain. Unsupervised learning of finite mixtures. *IEEE Transactions on PAMI*, 24(3) :381–396, 2002.

[9] Chris Fraley and Adrian E. Raftery. How many clusters ? Which clustering method ? Answers via model-based cluster analysis. *The Computer Journal*, 41(8) :578–588, 1998.

[10] U. Gargi, Y. Deng, and D. R. Tretter. Managing and searching personal photo collections. Technical Report HPL-2002-67, HP Laboratories, Palo Alto, Mar. 2002.

[11] M. Gelgon and K. Tilhou. Automated multimedia diaries of mobile dev ice users needs summarization. In *4th Int. Symp. on Mobile HCI'2002*, Lecture Notes in Computer Science, Pisa, Italy, sep 2002.

[12] M. Gelgon and K. Tilhou. Structuring the personal multimedia collection of a mobile device user based on geolocation. In *IEEE Int. conf. on Multimedia and Expo (ICME'2002)*, pages 248–252, Lausanne, Switzerland, aug. 2002.

[13] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *ACM Joint Conference on Digital Libraries JDCL*, Jun. 2002.

[14] P. Gros, R. Fablet, and P. Bouthemy. *New descriptors for image and video indexing*, chapter in State-of-the-Art in Content-Based Image and Video Retrieval, H. Burkhardt, H.-P. Kriegel, R. Veltkamp (eds). Kluwer, 2001.

[15] H. Kang and B. Shneiderman. Visualization methods for personal photo collections : Browsing and searching in the photofinder. In *IEEE Int. Conf. on Multimedia and Expo ICME'2000 (III)*, pages 1539–1542, New York, USA, 2000.

[16] A. Loui and A. E. Savakis. Automatic image event segmentation and quality screening for albuming applications. In *IEEE Proceedings Int. Conf. on Multimedia and Expo (ICME'2000)*, pages 1125–1128, New York, USA, aug. 2000.

[17] N. Marmasse and C. Schmandt. Location-aware information delivery with commotion. In *Handheld and Ubiquitous Computing HUC'2000, Second Int. Symp.*, pages 157–171, Bristol, UK, Sep. 25–27 2000.

[18] T. J. Mills, D. Pye, D. Sinclair, and K. R. Wood. Shoebox : A digital photo management system. Technical Report MSR 2000.10, AT&T Laboratories Cambridge, Cambridge, England, 2000.

[19] Y. Neuvo and J. Yrjanainen. Wireless meets multimedia : new products and services. In *IEEE Int. Conf on Image Processing (ICIP'02)*, pages I : 1–4, Rochester, USA, sep. 2002.

[20] J. C. Platt and B. A. Field M. Czerwinski. Phototoc : Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research, Feb. 2002.

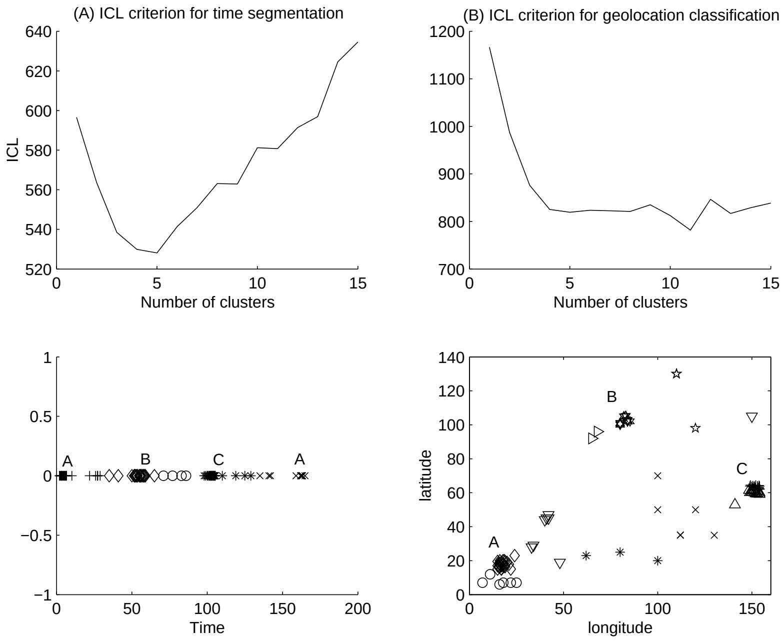
[21] K. Rodden. How do people organise their photographs ? In *21st Annual Colloquium on Information Retrieval Research*, Glasgow, Apr. 1999.

[22] A. E. Savakis, S. C. Etz, and A. Loui. Evaluation of image appeal in photographic images. In *SPIE Conf. on Human Vision and Electronic Imaging V*, jan. 2000.

[23] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12) :1349–1380, dec. 2000.

[24] A. Soronen and V. Tuomisto. Mobile image messaging - anticipating the outlines of the usage culture. In *4th Int. Symp. on Mobile HCI'2002*, pages 359–363, Pisa, Italy, sep 2002.

[25] Y. Zhao. Standardization of mobile phone positioning for 3G systems. *IEEE Communications Magazine*, pages 108–116, Jul. 2002.




4

**Fig. 2**. Time-based and geolocation-based classification of pictures : (top row) optimal values of the ICL criterion obtained for each candidate model complexity. (bottom row) classification obtained for the optimal model complexity, as found above.

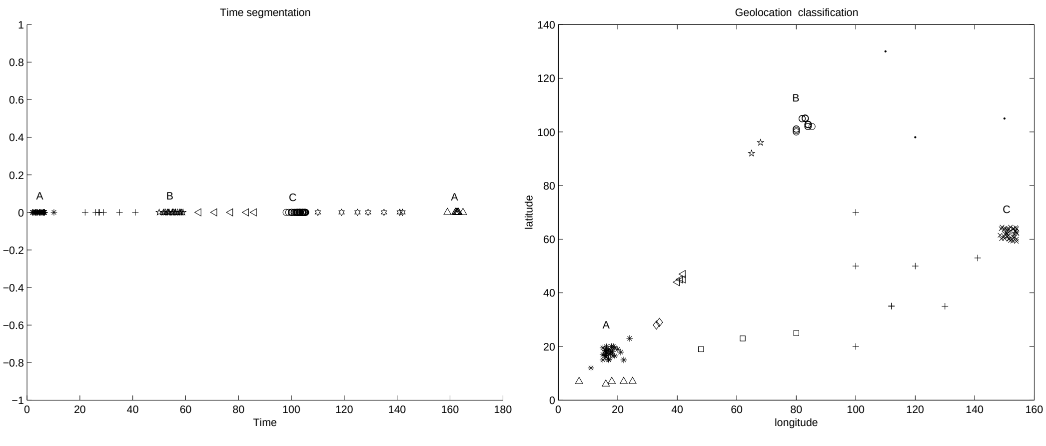



**Fig. 3**. Manually-defined “ground truth” classifications in time and space.

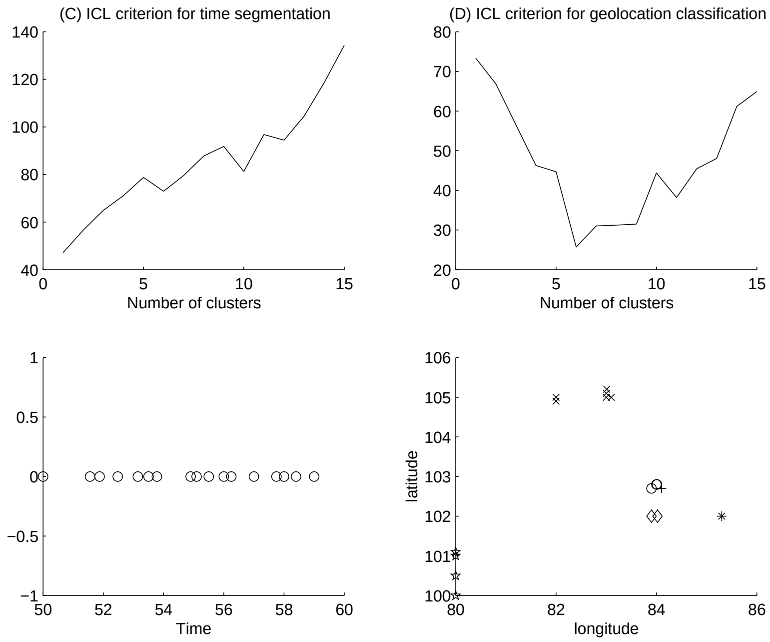



**Fig. 4**. On are “B”, time-based and geolocation-based classification of pictures : (top row) optimal values of the ICL criterion obtained for each candidate model complexity. (bottom row) classification obtained for the optimal model complexity, as found above.

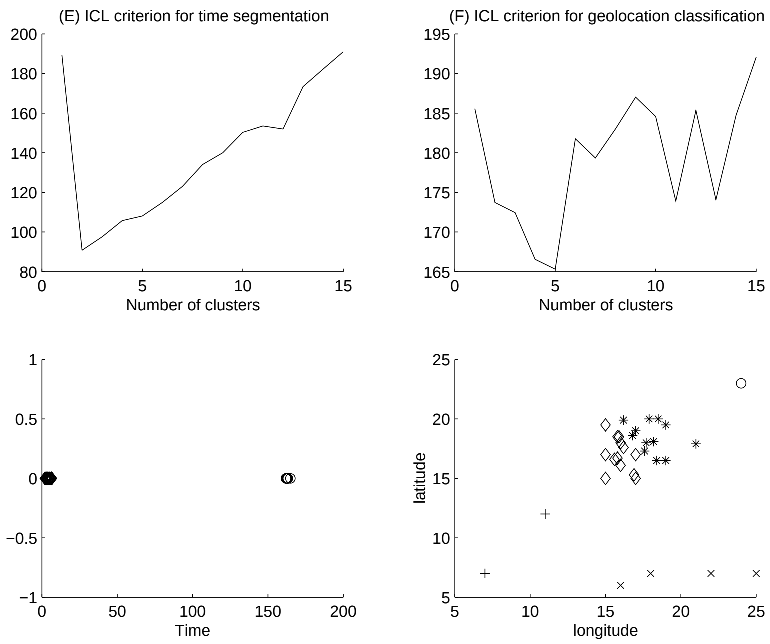



**Fig. 5**. On are “A”, time-based and geolocation-based classification of pictures : (top row) optimal values of the ICL criterion obtained for each candidate model complexity. (bottom row) classification obtained for the optimal model complexity, as found above.